Macromolecular crystallography with synchrotron radiation

JOHN R. HELLIWELL
Professor of Structural Chemistry
at the
University of Manchester
and
Joint Appointee
at
SERC, Daresbury Laboratory

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York NY 10011–4211, USA
10 Stamford Road, Oakleigh, Victoria 3166, Australia

First published 1992

Printed in Great Britain
at the University Press, Cambridge

A catalogue record of this book is available from the British Library

Library of Congress cataloguing in publication data

Helliwell, John R.
Macromolecular crystallography with synchrotron radiation / John R. Helliwell.
p. cm.
Includes bibliographical references and index.
ISBN 0–521–33467–5 (hardcover)
1. X-ray crystallography. 2. Synchrotron radiation. 3. Proteins—Analysis. 4. Nucleic acids—Analysis. 5. Viruses. I. Title.
QP519.9.X72H45 1992
547.7'046—dc20 91–10541 CIP

ISBN 0 521 33467 5 hardback

Contents

CHAPTER 1

Introduction

Macromolecular crystallography is a very powerful method used to study complex biological systems. The structures of a wide variety of proteins, nucleic acids and their assemblies have been determined at atomic or near-atomic resolution. As a result, a detailed understanding has been gained of various living processes such as enzyme catalysis, the immune response, the encoding of hereditary information, viral infection and photosynthesis.

The first X-ray diffraction photograph ever taken was from copper sulphate by Friedrich and Knipping at von Laue's suggestion in 1912. In the following year W. L. Bragg deduced the crystal structure of sodium chloride from Laue photographs. A variety of relatively small molecular structures were then solved at an increasing rate.

The first X-ray diffraction pictures of a protein crystal were taken in 1934 by Bernal in Cambridge, but in those days the data quality was crude and the techniques for deriving a crystal structure of a macromolecule from the X-ray data were not sufficiently developed. The advent of the computer has been a critical development.

The first protein structures to be determined were myoglobin and haemoglobin in the late 1950s by Kendrew *et al* (1958) and Perutz *et al* (1960). From then on a steadily increasing number of protein structures have become known. Nowadays, once a suitable crystal is available, a new structure of a protein or even a virus can, in favourable circumstances, be determined in a year or less. In the case where there is a closely related structure available then a new crystal structure may be obtained in as little as 1–2 weeks.

Crystallographic techniques are facilitated to a considerable extent by the degree of sophistication of the technology used at the various stages

leading to a macromolecular crystal structure determination. These stages are:

crystallisation of the pure sample;
determination of crystal unit cell parameters;
X-ray data collection to a given resolution;
solution of the crystallographic phase problem;
interpretation of the electron density map;
refinement of the molecular model against the observed data.

Detailing the function of the molecule then additionally involves substrate or inhibitor (drug) binding in the cases of enzymes or viruses and/or site directed mutagenesis of a protein and subsequent X-ray analysis. Sophisticated technologies are used which include:

genetic engineering to improve protein preparation and purification;
robotic machines to automate crystallisation procedures;
synchrotrons to provide intense, collimated and tunable X-rays to deal with problems associated with small samples, dense diffraction patterns (large unit cells), weak scattering, radiation damage and the phase problem, and to allow time resolved studies and the measurement of diffuse scattering;
electronic detectors and novel X-ray sensitive materials such as image plates and storage phosphors coupled to powerful computer workstations to improve the efficiency with which data are recorded and processed;
advanced computer graphics for molecular modelling and interpretation;
parallel and vector processor computers for refinement of the large numbers of parameters against the available data;
computer data bases and expert systems to study the relationships between protein structures.

This huge investment of skills and technologies is indicative of the paramount importance of X-ray crystallography in unravelling the detailed mechanisms of life.

1.1 WHY DO WE WANT TO KNOW A MACROMOLECULAR STRUCTURE?

The processes of life depend fundamentally on the atomic and geometrical structure and interactions of the molecules in the living cell. Thus molecular biology is fascinating because knowledge of the structure and action of these molecules gives a clue to the understanding of life. Enzymes, for example, are nature's catalysts. Nearly all reactions in the living cell occur at the right moment and with sufficient speed because of the catalytic action of an enzyme. They are extremely efficient, highly selective and orders of magnitude better than inorganic and organo-metallic catalysts. Most enzymes are proteins and composed of one or more polypeptide chains constructed from the 20 naturally occurring amino acids. The chains are folded into a globular shape and this three-dimensional structure determines the chemical and physical behaviour and the resulting function of the protein.

To determine these complicated structures the only general method available is X-ray diffraction of the single crystals of these materials. Although the structure of small proteins (molecular weight (MW) less than about 10000 daltons (D)) can be determined in solution with nuclear magnetic resonance (NMR) spectroscopy and the assembly of proteins in a complex can be studied with electron microscopy, only X-ray diffraction can give the three-dimensional structure of small as well as large proteins with a precision of about 0.1–0.2 Å.

Because most macromolecules exert their biological action in an aqueous solution, one may ask whether the molecular structure of a macromolecule in the crystal is a fair representation of the protein structure in solution. Since enzymes can be fully active in the crystalline state the answer is obviously that it is.

A macromolecular crystal usually consists of between 30% and 80% solvent of crystallisation. Hence, the enzyme active site, for example, can be accessible to these solvent channels and able therefore to catalyse conversion of a reactant to product. Because of this observation alone one is able to say that the crystal structure is directly relevant in helping to determine the macromolecule's functional state. Of course, the results themselves, defining the structure, do make chemical sense.

The greatest variety of biological structures is within proteins and much of the structural work in biology has been devoted to proteins. Also, the structures of a number of t-RNAs have been solved; a specific

t-RNA transports an amino acid to the protein synthetic machinery for translation into a growing polypeptide chain. More recently the structures of a number of DNA-binding proteins have been determined and from this work general principles for the interaction between DNA and proteins have been derived (for a review, see Steitz (1990)).

X-ray crystallography has also been used to unravel the structure of various enormous virus particles. Viruses have a protein coat which envelops the nucleic acid. The function of this coat is to protect the nucleic acid. In higher organisms antibodies are generated against this coating as a defence mechanism. From its three-dimensional structure one can see which parts of the protein coat are at the surface of the virus particle and which therefore are the potential sites for the generation of antibodies. These form the basis of ideas for the design of new vaccines.

Synchrotron radiation (SR) has been used to determine the structure of a variety of virus structures. The determination of the rhinovirus and the mengo virus structures has revealed an interesting surface morphology comprising a 'canyon'. The canyon has been shown to accommodate certain drug molecules, the effect of which is to prevent the virus from injecting its nucleic acid into the host. It may prove feasible therefore to treat certain viral infections directly with drugs, providing a major breakthrough in the treatment of active viral infections. Subsequently the structure of the foot and mouth disease virus has been determined. Although the canyon appears not to be present in this structure, it is expected that new forms of treatment for this disease will be forthcoming as a result of the knowledge of the structure.

The need for the three-dimensional structure of an enzyme to be known in order to understand its catalytic action has been emphasised. However, the structure is also indispensable for two other purposes:

The rational design of drugs for inhibiting the action of enzymes related to illnesses (e.g. penicillin destroying enzymes in bacteria): this can replace the more conventional, time-consuming screening of enormous numbers of compounds as the usual method for drug design.

Protein engineering: a development of great technical importance is the application of genetic engineering for the production of modified proteins with improved properties. Modification should be of the right kind in the right position (i.e. site specific) and this can only be

derived from the three-dimensional structure of the protein.

Hence, we can say that molecular structures of biological materials are needed for several purposes:

a complete understanding of the processes in the living cell, e.g. the action of enzymes, or the transfer of genetic information;
the rational design of drugs;
the modification of proteins via genetic engineering;
vaccine design.

X-ray crystallography is the most versatile method for determining these structures in detail, provided that suitable crystals can be grown.

1.2 THE IMPORTANCE OF SR IN MACROMOLECULAR CRYSTALLOGRAPHY

Particle accelerators were originally developed for high energy physics research into the subatomic structure of matter. The SR, which was produced by accelerators in that context, was a nuisance by-product – an energy loss process. The early stages of the utilisation of SR were therefore parasitic on the high energy physics machines whose parameters were, of course, not optimised for SR. However, SR has become well recognised in its own right as a major research tool in biology, chemistry and physics. Particle accelerators began to be designed as dedicated to SR production with parameters optimised solely for this work, e.g. with long lifetime beams and stable source positions; the Daresbury Synchrotron Radiation Source (SRS) was the first dedicated high energy source, which came on-line in 1981 (figure 1.1).

The technology of particle accelerators designed for SR production has advanced considerably so that it is now possible to induce sophisticated particle trajectories in special magnets known as wigglers and undulators (table 1.1). These new types of magnets can extend both the available spectral range of emitted photons as well as the brightness and brilliance compared with radiation from charged particles in simple circular trajectories in ordinary bending magnets. These enhancements from undulator and multipole wigglers at low emittance sources, compared with a bending magnet, cover five orders of magnitude. The emission from a bending magnet is, in itself, already two orders of magnitude higher than a laboratory X-ray, emission line source.

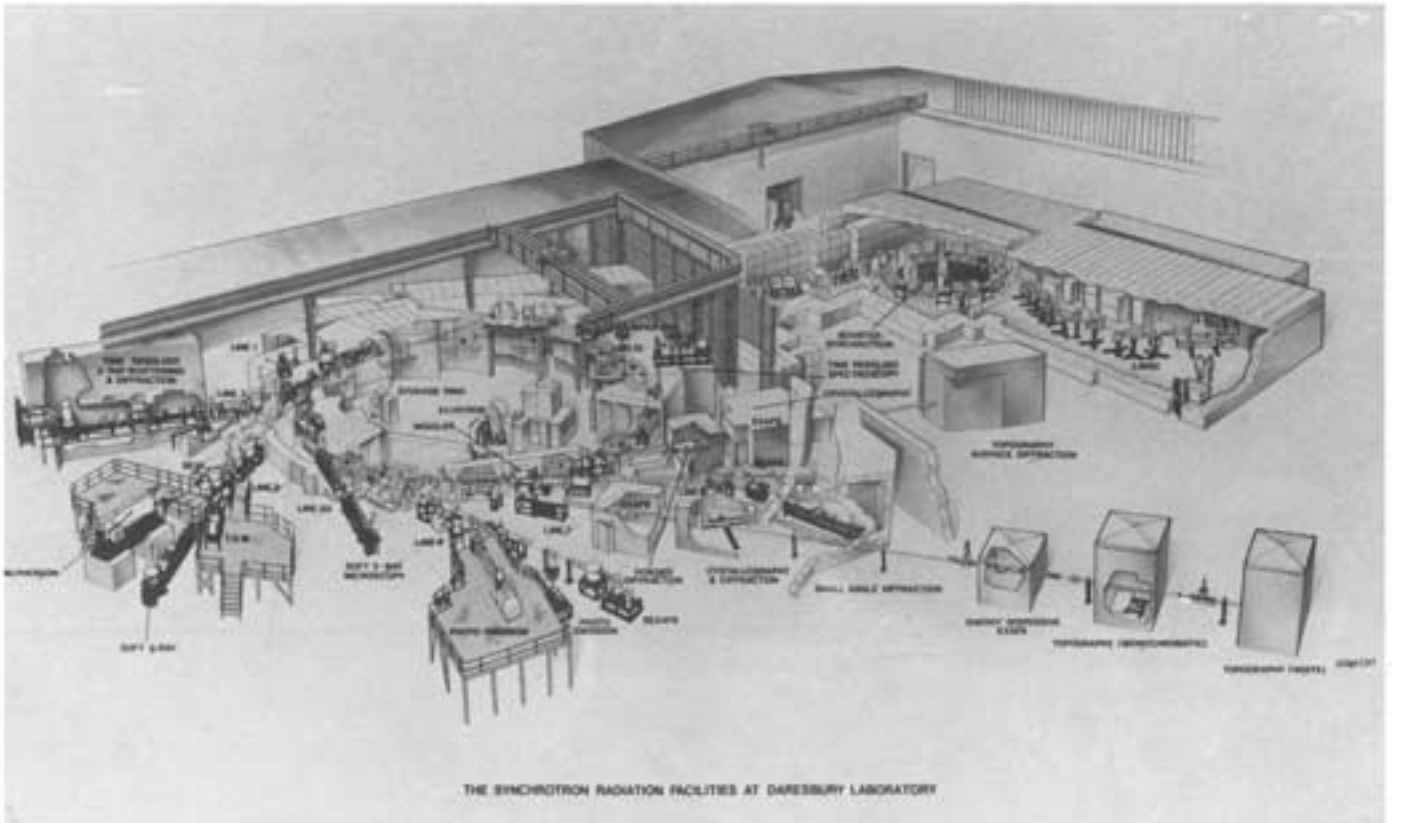

Figure 1.1 A layout of the SRS at Daresbury and its beam lines and workstations. The two labelled crystallography are for macromolecular crystal X-ray data collection. Obviously many other techniques are also housed at SR sources as is evident from this figure. Reproduced with the permission of SERC, Daresbury Laboratory.

Table 1.1. *The first eighteen beam lines planned at the ESRF Grenoble (from Miller (1990) reproduced with permission). The ESRF overall layout is shown in Figure 4.7 (in chapter 4).*

Beam line	Scientific areas	Source
1. Microfocus	Microdiffraction Small angle scattering High pressure	Undulator 0.8–3.0 Å
2. Multipole wiggler/ materials science	Small molecule crystallography Magnetic scattering	Wiggler 4–60 keV
3. Multipole wiggler/ biology	Macromolecular crystallography Laue and monochromatic modes High pressure energy dispersive studies	Wiggler 4–60 keV
4. High flux beam line	Real time small angle scattering Monochromatic protein crystallography	Undulator, λ tunable around 1.0 Å
5. High energy X-ray scattering	Gamma-ray diffraction Small angle scattering Compton scattering	Multipole wiggler or wavelength shifter
6. Circular polarization	Dichroism in EXAFS, SEXAFS Spin-dependent photoemission Microscopy at 2.5 keV	Helical undulator $E \leqslant 4$ keV
7. Surface diffraction	Surface structural studies Phase transitions Growth mechanisms Liquid surface diffraction	Undulator $K_{max} = 1.85$
8. Dispersive EXAFS	Time resolved structural studies	Wiggler, or tapered undulator
9. Undulator, open beam line		
10. Bending magnet open beam line		
11. Mössbauer/ High resolution Inelastic	Nuclear Bragg scattering High resolution (5–100 meV) Inelastic scattering at 0–5 eV energy transfer; electronic and vibrational excitations	Undulator ~ 14 keV
12. Asymmetric wiggler beam line	Magnetic scattering	Asymmetric wiggler
13. Surface science	SEXAFS and standing waves techniques	Undulator
14. High energy wiggler	Microtomography; possibly angiography	Wiggler
15. Powder diffraction	Powder diffraction for structure determination	Bending magnet later undulator

Table 1.1. (*cont.*)

Beam line	Scientific areas	Source
16. Wiggler Long beam line (75 m)	Topography (possibly second Laue station)	Multipole wiggler
17. Anomalous scattering beam line	Anomalous scattering	Undulator
18. EXAFS	Two EXAFS stations	Bending magnet

The properties of SR can be summarised as:

(*a*) high flux;
(*b*) high brightness (well collimated);
(*c*) high brilliance (small source size and well collimated);
(*d*) tunable over a wide wavelength range (X-rays–infra-red);
(*e*) polarised;
(*f*) defined time structure;
(*g*) calculable spectra.

The special combination of properties of SR finds widespread application in macromolecular crystallography.

Weak scattering is a feature of these crystals. The atomic numbers of the elements constituting the crystal (mainly carbon, nitrogen, oxygen, hydrogen and sulphur) means that the scattering of X-rays is relatively weak. In addition, the large fraction of solvent further reduces the scattering into the Bragg reflections. The resolution limit is rather poor, indicative of extensive flexibility and/or disorder in the crystal. In order to maximise the information that is derived from these crystals it is necessary, on conventional X-ray sources, to measure the high resolution, high angle data with long counting times. Unfortunately, radiation damage to the sample degrades the high resolution data first. The high intensity and collimation of the SR beam reduces the exposure time required by several orders of magnitude. As a result it is observed that radiation damage to the sample is reduced, i.e. more reflections can be measured from a crystal with a short exposure and a high intensity than with a long exposure and a low intensity. Moreover, short wavelength SR beams reduce the fraction of absorbed X-rays and again the damage. The main alternative for beating radiation damage is to use sample

freezing. If successful the lifetime of such a sample is very long. Unfortunately, it is not always possible to do this and when successful it is almost always accompanied by an increase in mosaic spread (e.g. 0.02° increases up to 0.5°). This causes the smearing out of diffraction spots. In some instances of large unit cell studies, e.g. crystallography of ribosomes, freezing is essential to preserve the resolution limit (to ≈5 Å) and the SR beam is then used to reduce the time and therefore burden of data collection to an acceptable level.

Virus and ribosome crystallography is one of the main uses of SR and one which takes advantage of a combination of SR properties such as the high flux, the fine collimation and short wavelengths.

Often large crystals can be elusive and so the intense SR beam is used to compensate for small sample volumes. The high brightness of SR allows a tiny incident beam to be brought onto a small crystal cross section area thus minimising any extraneous matter in the beam and therefore producing a reasonable signal to noise in the diffraction pattern. Small crystals are also vulnerable to radiation damage in that there are fewer unit cells available to yield diffraction intensity data of the required statistical precision. Freezing of the sample is required to preserve the lifetime of a small sample in the beam.

Rapid data collection has opened up the possibility of doing time resolved macromolecular crystallography of, for example, enzymatic processes. That this is feasible at all relies on the large solvent content of these crystals which permits function in this solid state. It is possible now to measure accurate diffraction data with an exposure time of a second or so, capable of yielding molecular structure detail of the necessary precision and sensitivity (e.g. to see a single water molecule binding to a protein). Exposure times have also been realised as short as 120 ps using an undulator as source.

The solution of a new macromolecular structure by crystallography requires that the phases of individual reflections are determined to allow a Fourier synthesis to be calculated. The method of multiple isomorphous replacement (usually with anomalous scattering) is used extensively whereby intensity changes are induced. As a result the positions of the heavy atoms are determined either by Patterson or direct methods and then the phases calculated. Usually, but not always, it is possible to interpret the electron density map. The anomalous scattering phasing power of a heavy atom derivative can now be rather easily optimised by appropriate choice of wavelength of the SR beam. The absorption edges of these high atomic number derivative atoms occur in the range

0.6 Å$<\lambda<$1.1 Å. In this range the protein crystal absorption is virtually eliminated for a typical crystal $<$1 mm in dimension. Sometimes it is not possible to make heavy atom derivatives owing to the chemical nature of a specific protein or the particular crystal form. Many proteins contain an essential metal atom or alternatively selenium can be incorporated into a protein. Similarly bromine can be incorporated into a nucleotide. In all these cases data can be collected at multiple wavelengths using SR and this allows phases to be determined. Protein structures have now been solved by several variants of these methods. This is an important technical capability because it either reduces the number of heavy atom derivatives that need to be found for isomorphous replacement or allows phase determination from a single crystal.

Future applications of SR in macromolecular crystallography involve extrapolations from the use of existing SR sources to the next generation of sources such as the European Synchrotron Radiation Facility (ESRF) in Grenoble or the Advanced Photon Source (APS) at Argonne or the Super Photon Ring (SPRING) in Japan. Shorter exposure times in time resolved work will be realised and smaller sample volumes and larger unit cells studied. The sophistication of our technical understanding is also improving. For example, the polarisation properties of the beam have yet to be exploited in the use of the dichroism of the scattering of heavier atoms in phasing reflections. It may also become routine for the diffuse scattering, which is rather weak, to be measured and interpreted. There are also other possible uses of SR, e.g. magnetic scattering. Moreover, the use of ultra-short wavelength (0.33 Å) beams from X-ray undulators could well provide a leap forward in data accuracy and molecular structure determination.

SR has revolutionised experimental X-ray crystallography and there are other fascinating developments in store.